Architecture of bacterial replication initiation complexes: orisomes from four unrelated bacteria

Anna ZAWILAK-PAWLIK*[1], Agnieszka KOIS*[1], Jerzy MAJKA†, Dagmara JAKIMOWICZ*‡, Aleksandra SMULCZYK-KRAWCZYSZYN*, Walter MESSER§ and Jolanta ZAKRZEWSKA-CZERWIŃSKA*[2]

*Ludwik Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, ul. Weigla 12, 53-114 Wrocław, Poland, †Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, U.S.A., and ‡John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, U.K., and §Max-Planck-Institut für Molekulare Genetik, Berlin-Dahlem, Ihnenstrasse 73, D-14195 Germany

Bacterial chromosome replication is mediated by single initiator protein, DnaA, that interacts specifically with multiple DnaA boxes located within the origin (*oriC*). We compared the architecture of the DnaA–origin complexes of evolutionarily distantly related eubacteria: two Gram-negative organisms, *Escherichia coli* and *Helicobacter pylori*, and two Gram-positive organisms, *Mycobacterium tuberculosis* and *Streptomyces coelicolor*. Their origins vary in size (from approx. 200 to 1000 bp) and number of DnaA boxes (from 5 to 19). The results indicate that: (i) different DnaA proteins exhibit various affinities toward single DnaA boxes, (ii) spatial arrangement of two DnaA boxes is crucial for the *H. pylori* and *S. coelicolor* DnaA proteins, but not for *E. coli* and *M. tuberculosis* proteins, and (iii) the *oriC* regions are optimally adjusted to their cognate DnaA proteins. The primary functions of multiple DnaA boxes are to determine the positioning and order of assembly of the DnaA molecules. Gradual transition from the sequence-specific binding of the DnaA protein to binding through co-operative protein–protein interactions seems to be a common conserved strategy to generate oligomeric initiator complexes bound to multiple sites within the chromosomal, plasmid and virial origins.

Key words: DnaA box, DnaA protein, *oriC* region, orisome (protein–*oriC* complex).

INTRODUCTION

In eubacteria and eukaryotes, and very likely in archaea as well, replication is controlled at the initiation stage. The initiation of DNA replication starts with the binding of specific initiator protein(s) to DNA sites, termed origins, and results in the localized unwinding of the DNA duplex and the establishment of replication forks [1,2]. Chromosomal replication initiates from single (most eubacteria and archaea, e.g. *Pyrococcus abyssi*) [3], two (archaea, e.g. *Sulfolobus solfataricus*) [4] or multiple replication origins (eukaryotes) [5,6]. Recent studies suggest that the mechanism of initiation of DNA replication is similar across all domains of life [7]. Genes encoding initiator proteins and binding sites for these proteins have been identified in chromosomes of bacteria, archaea and eukaryotes, as well as in plasmids and viruses.

Many initiator proteins belong to the AAA+ (ATPase associated with a variety of cellular activities) family of ATPases; they are active in the ATP-bound form and inactivated by hydrolysis of ATP to ADP (for a review, see [8]). Bacterial chromosome replication is mediated by a single initiator protein, DnaA; numerous molecules of the DnaA protein assemble into large complexes at the *oriC* (for a review, see [9]). Recent structural studies demonstrated that DnaA comprises an AAA+ ATPase fold [10]. In contrast, the eukaryotic initiator ORC (origin recognition complex) is a six-subunit heteromultimer that binds the origin region (for reviews, see [5,6]). Three of the ORC proteins possess AAA+-type ATP-binding motifs [8]. In archaea, replication is initiated by Orc1/Cdc6 protein(s) that are homologous with the eukaryotic initiator proteins Orc1 and Cdc6 [8]. There is no significant sequence homology between bacterial and archaeal/eukaryotic initiator proteins, apart from residues involved in nucleotide binding. However, the spatial arrangement of the conserved motifs in the DnaA protein (homo-oligomerization, ATPase and DNA-binding domains) is similar to that of the eukaryotic-like archaeal replication initiation factor Orc1/Cdc6, suggesting that the mechanistic elements of origin processing are conserved across bacterial, archaeal and eukaryotic life domains [7,8].

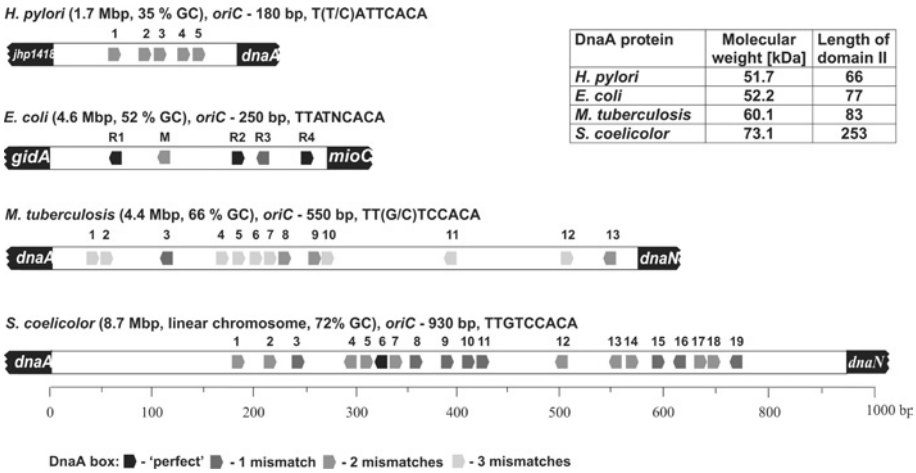
In eubacteria and archaea, the origins of replication are composed of modular elements, including sequences specifically bound by initiator proteins and AT-rich sequences that facilitate the opening of the double-stranded DNA helix. Some of these elements are multiply repeated, e.g. binding motifs for the eubacterial DnaA (DnaA boxes) and archaeal proteins Cdc6 (ORB, or origin recognition boxes) [4,7]. In eukaryotes, although the origins of replication are AT-rich, they share no detectable consensus sequences and therefore it is difficult to define distinct modular elements within their origins. Among eukaryotes, *Saccharomyces cerevisiae* seems to be an exception: its origin consists of an essential 11-base-pair consensus sequence (bound by ORC) and several additional sequence motifs that contribute to the initiation of replication [5,6].

The replication initiation process has been particularly well characterized in *Escherichia coli* [9,11]. The *E. coli* DnaA protein (52 kDa) binds to five non-palindromic nonamer sequences, the DnaA boxes that are localized within the *oriC*. Binding of 20–30 DnaA monomers promotes a local unwinding of an adjacent AT-rich region, whose single-stranded DNA is stabilized by binding of ATP–DnaA protein [9,12]. The unwound region provides an entry site for the DnaB–DnaC helicase complex,

Abbreviations used: AAA+, ATPase associated with a variety of cellular activities; *amp*^{r}, ampicillin resistance; DTT, dithiothreitol; EMSA, electrophoretic mobility-shift assay; HTH, helix–turn–helix; ORB, origin recognition boxes; ORC, origin recognition complex; *oriC*, origin of chromosomal replication; RU, resonance units; SPR, surface plasmon resonance; *tsr*^{r}, thiostrepton resistance.

[1] These authors contributed equally to this work.

[2] To whom correspondence should be addressed (email zakrzew@immuno.iitd.pan.wroc.pl).



DnaA protein	Molecular weight [kDa]	Length of domain II
H. pylori	51.7	66
E. coli	52.2	77
M. tuberculosis	60.1	83
S. coelicolor	73.1	253

Figure 1 Comparison of the *oriC* regions of the analysed bacteria

Genome size and GC content are given in parenthesis. For each analysed *oriC* region, its size and the consensus sequence for the DnaA box is presented. The insert presents features of the analysed DnaA proteins.

followed by other proteins required to form the replication fork [9].

Based on structural similarity and specific functions, four domains have been identified in DnaA [9,10,13]. The N-terminal domain I mediates protein–protein interactions, DnaA oligomerization and interaction with DnaB helicase [14,15]. Domain II appears to be less evolutionarily conserved and is probably a flexible linker that connects the N-terminal domain with the highly conserved domains III and IV. Domain III, besides the AAA+ ATPase motif, contains an additional oligomerization site. The C-terminal domain IV interacts specifically with DnaA boxes. Domain IV contains an HTH (helix–turn–helix) motif related to that found in the Trp repressor and many other DNA-binding proteins [16]. In addition to the HTH motif, two key functional elements of DnaA required for efficient DNA binding have been defined: (i) a DnaA signature sequence motif responsible for the affinity and specificity of DnaA box binding, and (ii) an extra basic loop that may contact the DNA minor groove or phosphate backbone [10].

The sequences of bacterial *oriC* regions are conserved only among closely related organisms. Bacterial replication origins vary in size (from approx. 200 to 1000 bp), although (nearly) all contain a few or more DnaA boxes and an AT-rich region [17].

Despite extensive work on the mechanism of initiation of DNA replication in prokaryotic and eukaryotic systems, several critical aspects of this mechanism and its control still remain obscure. One significant gap is a lack of information on the spatial arrangement of the initiation complex. An understanding of the biological and biochemical functions of the multiple initiator protein-binding sites required by a large group of chromosomal (DnaA boxes, ORB elements), plasmid (iterons), and viral (e.g. polyomavirus pentanucleotide sequences) replicons should contribute to solving this problem.

In the present paper, we compare the architecture of the DnaA–*oriC* complexes (orisomes) of evolutionarily distantly related eubacteria: two Gram-negative organisms, *E. coli* and *Helicobacter pylori*, and two Gram-positive organisms, *Mycobacterium tuberculosis* and *Streptomyces coelicolor*. The former two contain five DnaA boxes within their *oriC* regions, while the latter two possess numerous DnaA boxes: 13 and 19 in the *M. tuberculosis* and *S. coelicolor oriC* regions respectively. Thus these organisms provide a good and sufficiently differentiated group (see Figure 1) for comparative studies on the interactions between initiator protein and multiple cognate binding sites.

MATERIALS AND METHODS

DNA manipulations

E. coli XL1-Blue (*endA1*, *gyrA46*, *hsdR17*, *lac*, *recA1*, *supE44*, *thi*; F'*lac:lacI*^{q}, Δ(*lacZ*)*M15*, Tn*10*, *proA*$^+$, *proB*$^+$) served as the host for the plasmids. WM1785 (=W3110) and its *polA* derivative WM1838 (*polA*, *fadA*::Tnl0) were used as host strains in the *ori* assay [18]. Transformation of *S. coelicolor* has been described previously [19]. *S. coelicolor* transformants were selected for resistance to 10 μg of thiostrepton per ml. The DNA fragments and plasmids were purified using kits according to the manufacturer's protocols (Qiagen). DNA fragments for gel retardation were 5′-end-labelled using [γ-^{32}P]ATP and T4 polynucleotide kinase. Enzymes were supplied by Roche, Fermentas MBI and Gibco BRL. Isotopes were obtained from MP Biomedicals (Irvine, CA, U.S.A.). The oligonucleotides used for PCR or gel retardation were from Bionovo (Bioresearch Laboratory & Biochemicals, Legnica, Poland).

Plasmid and minichromosomes

The *oriC* regions of the analysed organisms were cloned into pBR322 plasmid. The *oriC* regions of *M. tuberculosis* and *S. coelicolor* were amplified by PCR using a pair of primers tailed by a BamHI motif and were cloned into pBR322 vector linearized by the same restriction enzyme. The resultant constructs were named pBR322_*MtoriC* and pBR322_*StoriC* respectively. The remainder constructs, pBR322_*HporiC* and pBR322_*EcoriC* (originally named pOC170) containing the *oriC* region from *H. pylori* and *E. coli* respectively were obtained as described previously [20,21].

For minichromosome replication studies in *S. coelicolor*, the PCR-amplified *oriC* regions were ligated with a DNA fragment carrying the *tsr*^{r} (thiostrepton resistance) gene and were then introduced into *S. coelicolor* protoplasts. Thiostrepton-resistant colonies were examined for the presence of minichromosomes (closed circular plasmid DNA) as described previously [22].

DnaA purification

Recombinant DnaA proteins of *E. coli*, *H. pylori*, *M. tuberculosis* and *S. coelicolor* were purified according to procedures described previously by Krause et al. [23], Zawilak et al. [24], Zawilak et al. [25] and Majka et al. [26] respectively. The purified DnaA proteins were more than 95 % homogeneous, as judged by SDS/PAGE analysis. ATP-bound DnaA proteins were used in all *in vitro* experiments.

Gel-retardation assay

For binding assays, ^{32}P-labelled DNA was incubated with DnaA protein in the presence of a non-specific competitor [poly-(dI-dC) · (dI-dC); Roche] at room temperature (20 °C) for 30 min in 0.5× Marians' binding buffer [1× Marians' buffer: 20 mM Hepes/KOH, pH 7.6, 5 mM magnesium acetate, 1 mM EDTA, 4 mM DTT (dithiothreitol), 0.2 % Triton X-100, 3 mM ATP and 50 μg/ml BSA] [27]. The bound complexes were separated by electrophoresis in 4 or 6 % polyacrylamide gels [0.25× TBE (89 mM Tris/89 mM borate/1 mM EDTA) at 4 V/cm, 4 °C]. Gels were dried and analysed by a Typhoon 8600 Variable Mode Imager and ImageQuant software. The apparent equilibrium dissociation constant K_d was determined using the method described by Carey [28]. A reaction mixture contained a fixed amount of DNA and various concentrations of DnaA protein. The DNA concentration chosen was much lower than the protein concentration required for half-maximal binding, so the protein concentration at half-maximal binding is very close to K_d (app). The K_d (app) was deduced from a curve [unbound DNA (%) against DnaA (nM)], based on the equation $K_d = [S] \cdot [P] \cdot [SP]^{-1}$, where [S] is DNA concentration, [P] is protein concentration and [SP] is DNA–protein complex concentration. When $[S] \ll K_d$, then $[P]_{free} \approx [P]_{total}$, so $K_d = [P]_{total} \cdot [S] \cdot [PS]^{-1}$.

SPR (surface plasmon resonance)

The BIAcore system (Biosensor AB) was used to study the interaction between protein and DNA. The system allows analysis of molecular interactions in real time by monitoring changes of the light angle inducing SPR. The SPR signal is directly proportional to the mass changes at the sensor surface and is expressed in RU (resonance units). For most proteins and DNA, 1000 RU corresponds to a surface concentration of approx. 1 ng/mm^2 (BIAevaluation software handbook, 1996, BIAcore AB). The biotinylated double-stranded oligonucleotides were immobilized on a streptavidin-coated SA chip of the BIAcore apparatus. Approx. 50 RU of DNA were immobilized. DNA loosely attached to the surface of the chip was removed with a pulse of 0.05 % SDS. In order to exclude the effects of mass transport on the kinetics of the protein–DNA interactions, the measurements were performed at various protein concentrations (0.39–100 nM) and at a continuous flow rate (100 μl · min^{-1}). The measurements were performed in KAC buffer (25 mM Hepes, 10 mM magnesium acetate, 100 potassium acetate, 5 mM EDTA and 0.005 % BIAcore surfactant P20) in the presence of poly(dI-dC) · (dI-dC) as a competitor. At the end of each cycle, bound DnaA protein was removed by washing with 0.05 % SDS. The kinetic constants for binding of DnaA protein to the double-stranded DNA were determined from the association and dissociation curves of the sensorgrams using the BIAevaluation version 3 program, as described previously [29].

SfiI probing

Plasmid pGEM®-T Easy-*Sfi*I*ScoriCSfi*I (150 ng) was incubated with different amounts of DnaA protein in binding buffer (10 mM Tris/HCl, pH 7.5, 10 mM $MgCl_2$, 50 mM NaCl and 1 mM DTT) at room temperature for 30 min. Then, SfiI digestion was carried out at 37 °C for 20 min according to the manufacturer's protocol (Amersham Biosciences). The reactions were stopped by incubation at 65 °C for 20 min. The SfiI cleavage products were separated on a 1 % agarose gel. After staining with Sybr Green I, the gel was analysed by a Typhoon 8600 Variable Mode Imager using ImageQuant software.

RESULTS

Different DnaA proteins exhibit various affinities toward single DnaA boxes

The consensus sequence for the *E. coli* DnaA box is 5′-TTATNCACA-3′. Binding studies revealed that the *E. coli* DnaA protein exhibits the highest affinity towards the 5′-TTATCCACA-3′ sequence (*Ec*DnaA box) that is named the 'perfect' DnaA box [9]. In high G + C content organisms, such as *Mycobacterium* and *Streptomyces*, the third position is replaced with G or C: 5′-TT-(G/C)TCCACA-3′ to give the *MtSc*DnaA box [30–32]. Comparison of DnaA boxes from the *H. pylori oriC* region allowed us to determine the consensus sequence: 5′-TTCTTCACA-3′ (*Hp*DnaA box) [33]. Since the consensus sequences for DnaA boxes differ slightly among the organisms analysed, it was interesting to compare the binding of the DnaA proteins to distinct single DnaA boxes (*Hp*DnaA, *Ec*DnaA and *MtSc*DnaA). To evaluate the interactions of the DnaA proteins with individual DnaA boxes, SPR and/or a gel-retardation assay was applied. Oligonucleotides containing a single DnaA box, *Hp*DnaA, *Ec*DnaA or *MtSc*DnaA box (Table 1), were end-labelled with biotin or [γ-^{32}P]ATP for SPR or a gel-retardation assay respectively and were then incubated with increasing amounts of a given DnaA protein. Representative binding experiments are shown in Figure 2. The K_d (app) values (Figure 2) were calculated from the sensorgrams and/or by quantification of gel-retardation assays as described previously in detail [29,34]. Our results show that the affinities of the analysed proteins for a single DnaA box vary significantly. The *E. coli* DnaA protein exhibited the highest affinity toward its own (*Ec*DnaA) box. This protein, in addition to its own DnaA box, also bound the *MtSc*DnaA box. As was shown previously, the *S. coelicolor* DnaA protein prefers the *E. coli* DnaA box over its own DnaA box [34]. The *H. pylori* DnaA bound only its own *Hp*DnaA box, although with weak affinity. What is more, no other analysed DnaA protein bound the *Hp*DnaA box. In contrast with the other proteins, the *M. tuberculosis* DnaA binds neither its cognate *MtSc*DnaA box nor other single DnaA boxes (Figure 2).

Spatial arrangement of two DnaA boxes is crucial for the *H. pylori* and *S. coelicolor* DnaA proteins, but not for *E. coli* or *M. tuberculosis* proteins

Our previous studies demonstrated that the *S. coelicolor* DnaA protein exhibits preference for DNA containing two DnaA boxes in reverse orientation separated by 3 bp; such a box arrangement has been found in the promoter region of the *dnaA* gene as well as in the *oriC* regions of different *Streptomyces* species, including *S. coelicolor*, e.g. the fifth and sixth DnaA boxes (Figure 1) [30,35,36]. The structure of the *H. pylori oriC* suggests that *H. pylori* DnaA may prefer the same orientation of two DnaA boxes (2 bp between them, *Hp* boxes-wt; Figure 1). Thus it seems interesting to examine the influence of orientation and distance between two DnaA boxes on the binding affinity of the analysed DnaA proteins. For this purpose, we synthesized a set of six oligonucleotides bearing two boxes in reverse (*Sc*boxes-wt, *Sc*boxes-wt + 10 bp, *Hp*boxes-rv) or direct (*Sc*boxes-fw,

Table 1 Oligonucelotides used in the present study

X, biotin.

Oligonucleotide	Sequence (5′ → 3′)	Used to prepare
*Hp*DnaAbox_fw	XTTTTAAGGCTTCATTCACATGTTATTCCT	Oligonucleotides for SPR and EMSA analysis
*Hp*DnaAbox_rv	AGGAATAACATGTGAATGAAGCCTTAAAA	
*Ec*DnaAbox_fw	XACAGAGTTATCCACAGTAGAT	
*Ec*DnaAbox_rv	ATCTACTGTGGATAACTCTGT	
*MtSc*DnaAbox_fw	XGAGACACTTGTCCACACAACT	
*MtSc*DnaAbox_rv	AGTTGTGTGGACAAGTGTCTC	
*Hp*boxes-wt_fw	XTTTTAAGGCTTCATTCACATTTCATTCACATGTTATTCCT	
*Hp*boxes-wt_rv	AGGAATAACATGTGAATGAAATGTGAATGAAGCCTTAAAA	
*Hp*boxes-rv_fw	XAAGGCTTCATTCACATTTGTGAATGATGTTAT	
*Hp*boxes-rv_rv	ATAACATCATTCACAAATGTGAATGAAGCCTT	
*Hp*boxes-wt + 10bp_fw	XAAGGCTTCATTCACATTAACCCTTAAATCATTCACATGTTAT	
*Hp*boxes-wt + 10bp_rv	ATAACATGTGAATGATTTAAGGGTTAATGTGAATGAAGCCTT	
*Sc*boxes-wt_fw	XAGACACTTGTCCACAGGCTGGGGACAACAACTT	
*Sc*boxes-wt_rv	AAGTTGTTGTCCCCAGCCTGTGGACAAGTGTCT	
*Sc*boxes-rv_fw	XAGACACTTGTCCACAGGCTTGTCCCCACAACTT	
*Sc*boxes-rv_rv	AAGTTGTGGGGACAAGCCTGTGGACAAGTGTCT	
*Sc*boxes-wt + 10bp_fw	XAGACACTTGTCCACAGACTAGTCCTGGCTGGGGACAACAACT	
*Sc*boxes-wt + 10bp_rv	AAGTTGTTGTCCCCAGCCAGGACTAGTCTGTGGACAAGTGTCT	
nonbox_fw	XATCAGTCACGTGATCAGATCA	
nonbox_rv	TGATCTGATCACGTGACTGAT	
*Hpori*_fw	CACATTCTCGTTAGATTAATCGC	*H. pylori oriC* region for EMSA
*Hpori*_rv	GCGTTAGGGTTGTATTTGAGTTG	
*Ecori*_fw	TGTGATCTCTTATTAGGATC	*E. coli oriC* region for EMSA
*Ecori*_rv	ACTCAAATAAGTATACAGATC	
*Mtori*_fw	CGGGATCCCACGGCGTGTTCTTCCGACAACG	*M. tuberculosis oriC* for EMSA
*Mtori*_rv	CGGGATCCTGCGCCCTTTCACCTCACGATGAG	
*ScoriSfi*Ibend_fw	GGATCCGGCCGTTACGGCCGTCCTCCACAG	*S. coelicolor oriC* for construction of pGEM®-T Easy-*Sfi*I*ScoriSfi*I
*Scori*bend_rv	CGGGATCCTCGTGACGAGGTGCGGTCGG	

*Hp*boxes-wt, *Hp*boxes-wt + 10 bp) orientation (see Figure 3). None of the *H. pylori* oligonucleotides was bound by the *S. coelicolor* DnaA (and vice versa; results not shown). Incubation of the *M. tuberculosis* or *E. coli* DnaA protein with each of the six oligonucleotides caused the appearance of single or two nucleoprotein complexes respectively; both proteins exhibited higher affinity toward the *S. coelicolor* boxes than toward the *H. pylori* DnaA boxes (results not shown). Neither the relative orientation nor the spacing between the two DnaA boxes significantly affected the affinity of the *E. coli* and *M. tuberculosis* DnaA for the DNA. However, if the *E. coli* DnaA binds as dimer, the orientation and distance of DnaA boxes does affect the binding affinity [11]. In contrast, the arrangement of boxes was important for the *H. pylori* DnaA protein, which exhibited the highest affinity toward the wild-type arrangement of DnaA boxes (*Hp*boxes-wt). The two *H. pylori* DnaA boxes in reverse orientation (*Hp*boxes-rv) were also bound by the *H. pylori* DnaA, albeit with lower affinity than the wild-type boxes. Addition of 10 bp (*Hp*boxes-wt + 10 bp) significantly reduced the interaction between *H. pylori* DnaA and two boxes (Figure 3). The *S. coelicolor* DnaA bound all three oligonucleotides; however, the boxes in direct orientation (*Sc*boxes-fw) were bound less efficiently. Separation of the DnaA boxes (*Sc*boxes-wt + 10 bp) did not change the affinity of the protein (mainly single nucleoprotein complex was formed).

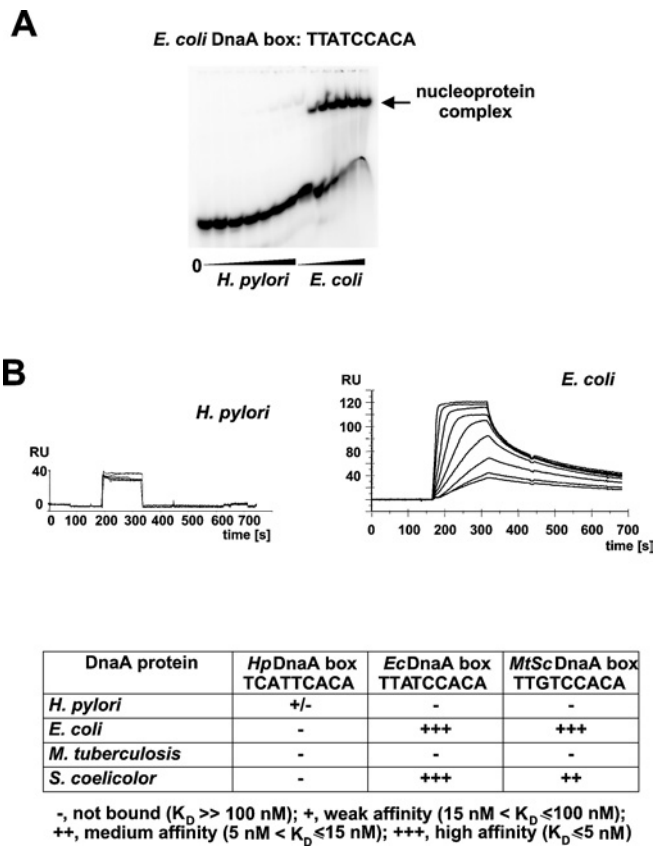
Thus, among the DnaA proteins analysed, the orientation of the boxes and distance between them have influence on the *H. pylori* and *S. coelicolor* DnaA binding.

The *oriC* regions are optimally adjusted to their cognate DnaA proteins

Interactions between whole *oriC* regions and DnaA proteins were analysed by a series of gel-retardation assays. For our studies, four *oriC* regions from *H. pylori*, *E. coli*, *M. tuberculosis* and *S. coelicolor* were amplified by PCR using the templates and pairs of primers listed in Table 1. The number of DnaA boxes within the *oriC* regions analysed varies from five (*H. pylori* and *E. coli*) to 19 (*S. coelicolor*) (Figure 1). The labelled *oriC* fragments were incubated with increasing amounts of purified DnaA proteins, and then nucleoprotein complexes were analysed in a 4 % native polyacrylamide gel. Formation of nucleoprotein complexes was studied in homologous (*oriC* and DnaA from the same organism) and heterologous (*oriC* and DnaA from two different organisms) systems. It is worthwhile to mention that DnaA proteins bind with the same affinity to linear and supercoiled *oriC* DNA ([21], and results not shown). DnaA binding to *oriC* is determined by the DNA sequence rather than by the topology. Therefore we used for our assay a linear form of DNA that allowed us to follow in an easy manner a nucleoprotein complex formation.

In homologous, as well as heterologous, systems, the nucleoprotein complexes observed were formed in a manner that is dependent on the protein concentration (Figure 4). In homologous systems, nucleoprotein complexes readily appeared at the lowest protein concentrations (Figure 4). In most heterologous systems (but not for *M. tuberculosis* DnaA), nucleoprotein complexes were detectable only at elevated protein concentrations. The *H. pylori* and *S. coelicolor* DnaA proteins and their *oriC* regions exhibited particularly low reciprocal affinity in heterologous system; diffuse nucleoprotein complexes were detectable only at the highest protein concentrations (Figure 4). In contrast, *M. tuberculosis* DnaA bound with a high avidity not only its own *oriC* region, but also ‘foreign’ *oriC* region from *E. coli* and *S. coelicolor* and, to some extent, *H. pylori oriC* (Figure 4).

Interestingly, interactions of the *E. coli* DnaA protein with each of the analysed *oriC* regions led to the formation of multiple discrete nucleoprotein complexes which were visible as a ladder



DnaA protein	*Hp*DnaA box TCATTCACA	*Ec*DnaA box TTATCCACA	*MtSc* DnaA box TTGTCCACA
H. pylori	+/-	-	-
E. coli	-	+++	+++
M. tuberculosis	-	-	-
S. coelicolor	-	+++	++

-, not bound (K_D >> 100 nM); +, weak affinity (15 nM < K_D ≤100 nM); ++, medium affinity (5 nM < K_D ≤15 nM); +++, high affinity (K_D ≤5 nM)

Figure 2 Interactions of DnaA proteins with a single DnaA box

(**A**) Gel retardation assay. The assay was performed using a ^{32}P-labelled double-stranded oligonucleotide containing the single *E. coli* DnaA box (*Ec*DnaAbox) (20 fmol). The DNA fragment was incubated with increasing amounts of *H. pylori* (tracks 2–8) or *E. coli* (tracks 9–14) DnaA. Tracks: 1, 0; 2 and 9, 1; 3, 5; 4 and 10, 10; 5 and 11, 25; 6 and 12, 50; 7 and 13, 100; 8 and 14, 250 nM DnaA. (**B**) SPR analysis. The top part shows the BIAcore sensorgram of the *H. pylori* or the *E. coli* DnaA protein binding to the *E. coli* DnaA box (*Ec*DnaAbox). The biotinylated double-stranded oligonucleotide was immobilized on a streptavidin-coated chip of the BIAcore apparatus and then incubated with increasing amounts of DnaA protein. The concentrations of the DnaA protein (from bottom to top) were, for *H. pylori*, 1, 5, 10, 25, 50, 100 and 500 nM, and, for *E. coli*, 1.2–300 nM in 2-fold increments. The bottom part shows the affinity of the DnaA proteins toward different types of single DnaA box. K_d (app) values were determined by gel retardation assay and/or by SPR as described in detail in [28,34]. The *M. tuberculosis* and *S. coelicolor* DnaA boxes have the same consensus sequence (TTGTCCACA).

of retarded bands (Figure 4). As the *E. coli* DnaA protein concentration increased, the complexity of the band pattern increased until a critical point was reached at which the ladder pattern disappeared and was replaced by diffuse highly retarded bands, indicative of large, but not very stable, complexes (Figure 4). The number of discrete retarded bands seemed to be proportional to the complexity of the *oriC* region analysed: the highest number of nucleoprotein complexes could be observed for the *S. coelicolor oriC* region that contains 19 DnaA boxes. The formation of discrete nucleoprotein complexes can be explained by monomeric binding of the *E. coli* DnaA to multiple DnaA boxes; at lower concentrations, the *E. coli* DnaA bound each of the DnaA boxes individually (discrete bands were visible on the electrophoretic mobility-shift assay gel; Figure 4), starting from high-affinity DnaA boxes (e.g. DnaA boxes R1 and R4; [21,37]).

H. pylori DnaA protein also bound DnaA boxes sequentially, but preferentially those from its own *oriC* region. *H. pylori* DnaA exhibited the highest affinity towards two closely spaced DnaA boxes that were arranged 'head–2 bp–tail'; such an arrangement is repeated within the *H. pylori oriC* region (Figure 1). The *H. pylori* DnaA protein bound randomly to the repeated pair of the DnaA boxes (**2**-**3** and **4**-**5**); each pair of DnaA boxes is probably bound by a dimer of the DnaA protein. Higher protein concentrations are required to observe binding to the DnaA box **1** that is not bound as an individual DnaA box [33].

In contrast with *E. coli* and *H. pylori,* incubation of the *M. tuberculosis* and *S. coelicolor* DnaA proteins with the *oriC* regions led to the formation of high-molecular-mass complexes already at a low protein concentration. *S. coelicolor* DnaA first bound the 'head–3 bp–head' DnaA boxes **5** and **6**, and then 'weaker' DnaA boxes [26]; the discrete complexes were only visible at the lowest protein concentration (Figure 4). For *M. tuberculosis* DnaA, discrete nucleoprotein complexes were not observed, even at a low protein concentration (Figure 4). The formation of higher-molecular-mass complexes, which were visible as a diffuse band(s) already at a low protein concentration, was probably caused by a property of the *M. tuberculosis* DnaA protein; it does not bind a single DnaA box and, for efficient binding, presumably requires at least two DnaA boxes, which are bound in a co-operative manner.

In order to elucidate further DnaA protein binding specificity with regard to the origin region, a series of competition gel-retardation assays were performed. All four labelled *oriC* fragments were incubated separately with each of the analysed DnaA proteins (Figure 5). Competition gel-retardation assays revealed that the DnaA proteins from *H. pylori* and *S. coelicolor* exhibited the highest affinity towards their own *oriC* regions (Figure 5). These experiments corroborated an earlier observation that the *oriC* regions of *H. pylori* and *S. coelicolor* were poorly recognized by heterologous DnaA proteins. The *M. tuberculosis* DnaA protein preferred the *oriC* region of *S. coelicolor* over its cognate origin: *S. coelicolor oriC* was completely bound at the lowest concentration of this protein (Figure 5).

It has been demonstrated that DnaA proteins mediated DNA bending of *oriC* regions [23,35,38]. The DnaA protein of *Bacillus subtilis* forms stable loops by interaction of protein molecules bound to the DnaA box groups separated by the long spacer, *dnaA* gene (~1400 bp). Interestingly, the *E. coli* DnaA is also able to loop out the *B. subtilis oriC* region [23]. The *oriC* region from *S. coelicolor* also contains two clusters of DnaA boxes; however, they are separated by a short spacer (120 bp; Figure 1). In order to check whether the *S. coelicolor oriC* is exclusively looped out by its cognate DnaA, we examined the ability of DnaA proteins from *E. coli* and *M. tuberculosis* to form the loop structure. For this purpose, the SfiI probing assay was applied. The interaction with two SfiI sites *in cis* involves the formation of a DNA loop between the sites, presumably as a result of the tetrameric protein binding simultaneously to both sites [39]. For SfiI probing, the *oriC* region of *S. coelicolor* was PCR-amplified with a pair of primers (*Scori**Sfi*Ibend_fw and *Scori*bend_rv) in such a way that two SfiI cleavage sites flank the *oriC*. The PCR product was cloned into a T-vector (pGEM®-T Easy), which does not contain any SfiI sites. The resulting plasmid pGEM®-T Easy-*Sfi*I*ScoriCSfi*I was incubated with different amounts of the *S. coelicolor*, *E. coli* or *M. tuberculosis* DnaA protein and then subjected to the SfiI enzyme cleavage. Under the conditions used, the SfiI digestion was inefficient, but was enhanced significantly (~3.0 times; Figure 6) by the *S. coelicolor* DnaA at a protein/DnaA box ratio of 1:1. In contrast, when *E. coli* or *M. tuberculosis* DnaA was incubated with pGEM®-T Easy-*Sfi*I*ScoriCSfi*I, no distinct enhancement of SfiI cleavage was observed (Figure 6). Thus SfiI probing demonstrated that the *S. coelicolor oriC* is exclusively bent by its own DnaA protein.

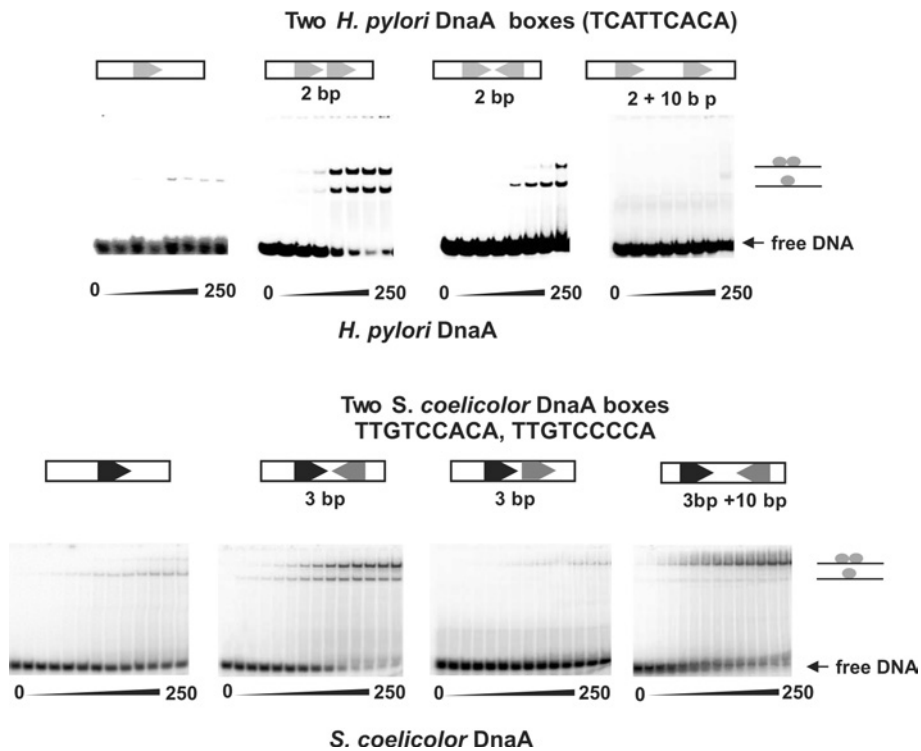


Figure 3 Interaction of DnaA proteins with two DnaA boxes

Gel retardation assay was performed using ^{32}P-labelled double-stranded oligonucleotides containing two DnaA boxes (10 fmol): *Hp*boxes-wt, *Hp*boxes-rv, *Hp*boxes-wt + 10bp, *Sc*boxes-wt, *Sc*boxes-rv and *Sc*boxes-wt + 10bp (Table 1). DNA fragments were incubated with increasing amounts of the *H. pylori* or *S. coelicolor* DnaA protein. The DNA–protein complexes were separated on a 6 % polyacrylamide gel. The concentrations of the DnaA protein were, for *H. pylori*, 1.95–250 nM and, for *S. coelicolor*, 0.06–250 nM in 2-fold increments.

Host specificity of *oriC* regions

In order to investigate the host specificity of the analysed *oriC* regions, we performed a set of heterologous transformations; the ability of the origins to promote replication was examined in *E. coli* and *S. coelicolor* (Table 2). For assay in *E. coli*, the PCR-amplified *oriC* regions were cloned into the pBR322 plasmid. The resultant constructs, named pBR322_*HporiC*, pBR322_*EcoriC*, pBR322_*MtoriC* and pBR322_*ScoriC*, were assayed for *oriC*-dependent initiation of replication *in vivo* using the frequently used *polA* system [40]. ColE1-type plasmids (such as pBR322) require DNA polymerase I for their replication. Thus only the constructs containing a functional *oriC* region are able to replicate (and confer pBR322-encoded ampicillin resistance, *amp*^{r}) in the absence of DNA polymerase I (in *E. coli polA*$^-$ strain). Similar to pBR322 itself (a negative control), none of the 'foreign' *oriC* regions gave ampicillin-resistant transformants in *polA* deficient strain; the *amp*^{r}-positive colonies did not appear even after prolonged incubation (of a few days). The pBR322_*EcoriC* was the only construct in which *oriC* was active (positive control). For a replication assay in *S. coelicolor*, minichromosomes (or shuttle vectors, containing, in addition to the *oriC* region and *tsr*^{r}, pMB1 replicon and *amp*^{r} to propagate in *E. coli*; results not shown) were constructed, and their ability to replicate *in vivo* was tested as described previously [22]. In contrast with p*ScoriC* minichromosome (positive control), p*EcoriC* minichromosome was found not to replicate in *S. coelicolor* (Table 2). Interestingly, despite the fact that ScDnaA protein binds the *oriC* region of *M. tuberculosis* with a high affinity, the *M. tuberculosis oriC* was not able to initiate minichromosome replication in *S. coelicolor*. Replication ability of the *H. pylori oriC* region in *S. coelicolor* was not tested, since this region exhibits particularly low affinity for the *Sc*DnaA protein.

DISCUSSION

Initiators of chromosomal replication are either single proteins (DnaA) or multisubunit complexes (six proteins; Orc1–6) that bind single or repetitive sites (e.g. DnaA boxes). One of the critical points of the initiation of replication is the formation of an initiation complex that promotes helical instability, unwinding and subsequent helicase loading. In eubacteria, the arrangement of DnaA boxes and their number are not conserved among eubacteria; *oriC* regions contain from a few to up to nearly 20 DnaA boxes, separated by various spacers [17]. In the present paper, we compared the similarities and differences in the formation of initiation complexes among four phylogenetically unrelated eubacteria: *H. pylori*, *E. coli*, *M. tuberculosis* and *S. coelicolor*; their origins contain different numbers of DnaA boxes (for details, see Figure 1).

The DnaA proteins analysed differ in binding affinity

The DnaA proteins analysed are able to bind a single DnaA box, with the exception of the *M. tuberculosis* DnaA protein. However, they exhibit different affinity towards single DnaA boxes; the *E. coli* and *H. pylori* DnaA proteins bind their cognate DnaA boxes with the highest and lowest affinity respectively. In contrast with the *E. coli* and *H. pylori* DnaA proteins, the *S. coelicolor* DnaA protein reveals higher affinity for the *E. coli* DnaA box than for its own DnaA box. Recently, analysis of the crystal structure

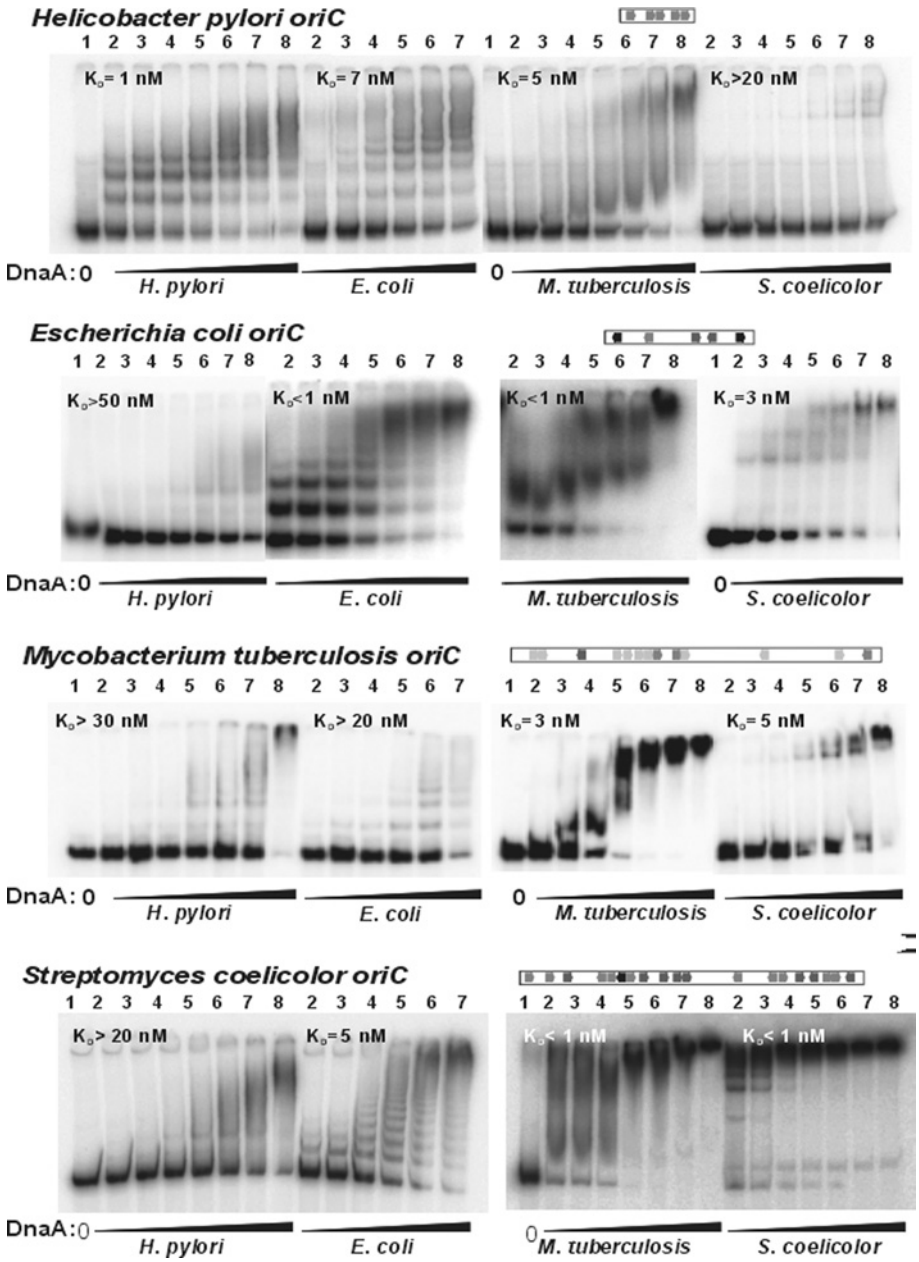


Figure 4 Interaction of DnaA proteins with *oriC* regions

Gel-retardation assay was pre-formed using ^{32}P-labelled *oriC* fragments (2–5 fmol) that were incubated with increasing amounts of the *H. pylori*, *E. coli*, *M. tuberculosis* or *S. coelicolor* DnaA protein. The DNA–protein complexes were separated on a 4 % polyacrylamide gel. The structure of the *oriC* regions (for details, see Figure 1) are shown above the gels. Tracks: 1, 0; 2, 1; 3, 2.5; 4, 5; 5, 10; 6, 25; 7, 50; 8, 100 nM DnaA. K_d (app) was determined using the method described by Carey [28] (see the Materials and methods section).

of the *E. coli* DnaA binding domain complexed with R1 DnaA box revealed that 18 amino acid residues have direct contact with DNA (see Figure 7): 14 residues interact with DNA backbone phosphate groups, four residues are involved in the base-specific interactions (two of them, Asp^{434} and Arg^{399} interact with both bases and phosphor groups) [16]. Sequence comparison of the *E. coli* binding domain with the corresponding protein fragments form *H. pylori*, *M. tuberculosis* and *S. coelicolor* showed principal differences that are summarized in Figure 7. First, interesting differences between the *E. coli* and *M. tuberculosis* or *H. pylori* DnaA binding domains concern residues involved directly in the base recognition: His^{439} and Pro^{423} both make van der Waals contact with the C-5 methyl group of T4 and T9* (*, opposite strand) respectively (Figure 7). His^{439} is replaced in the *M. tuberculosis* DnaA (and other mycobacterial DnaA proteins as well) by the non-closely related amino acid, tyrosine (Figure 7). In *E. coli*, His^{439} belongs to the A-signature motif that has been shown by extensive mutational analyses to define both the affinity and specificity of DnaA box binding. Thus it is tempting to conclude that the substitution (H439Y) may cause a decrease in the *M. tuberculosis* DnaA affinity. In the *H. pylori* DNA-binding domain, Pro^{423} (a well-conserved residue in other DnaAs) is replaced by leucine; this substitution probably causes reduction of protein affinity. It should be noted that the binding domain

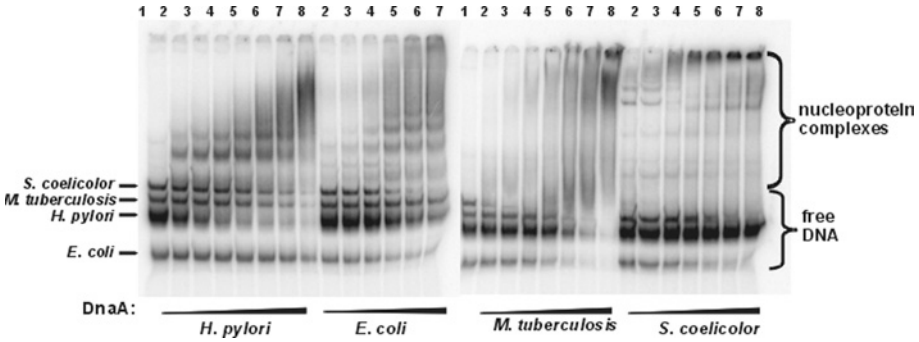


Figure 5 Competition gel retardation assay

All four ^{32}P-labelled *oriC* fragments (~2 fmol of each *oriC* fragment) were incubated together with increasing amounts of the *H. pylori*, *E. coli*, *M. tuberculosis* or *S. coelicolor* DnaA protein. Tracks: 1, 0; 2, 1; 3, 2.5; 4, 5; 5, 10; 6, 25; 7, 50; 8, 100 nM DnaA. The *H. pylori oriC* DNA fragment used for this analysis was longer (360 bp) than the *oriC* region of *H. pylori* (the fragment does not contain any DnaA box outside the *oriC* region, therefore the affinity of the HpDnaA protein towards the *oriC* region was not changed [20]).

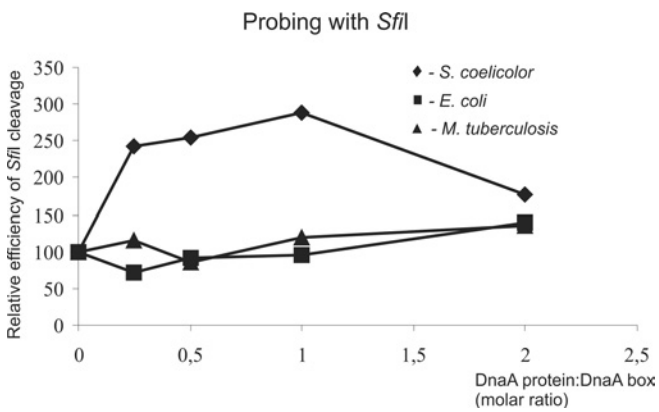


Figure 6 Bending of the *S. coelicolor oriC* region by DnaA proteins – probing of DnaA–*oriC* complexes with SfiI

Plasmid pGEM®-T Easy-*Sfi*I*Sco*ri*CSfi*I containing the *S. coelicolor oriC* region flanked by SfiI restriction sites was incubated with different amounts of the *E. coli*, the *M. tuberculosis* or the *S. coelicolor* DnaA protein and then subjected to SfiI enzyme cleavage. The SfiI enzyme digestion activity in the absence of DnaA protein was set at 100.

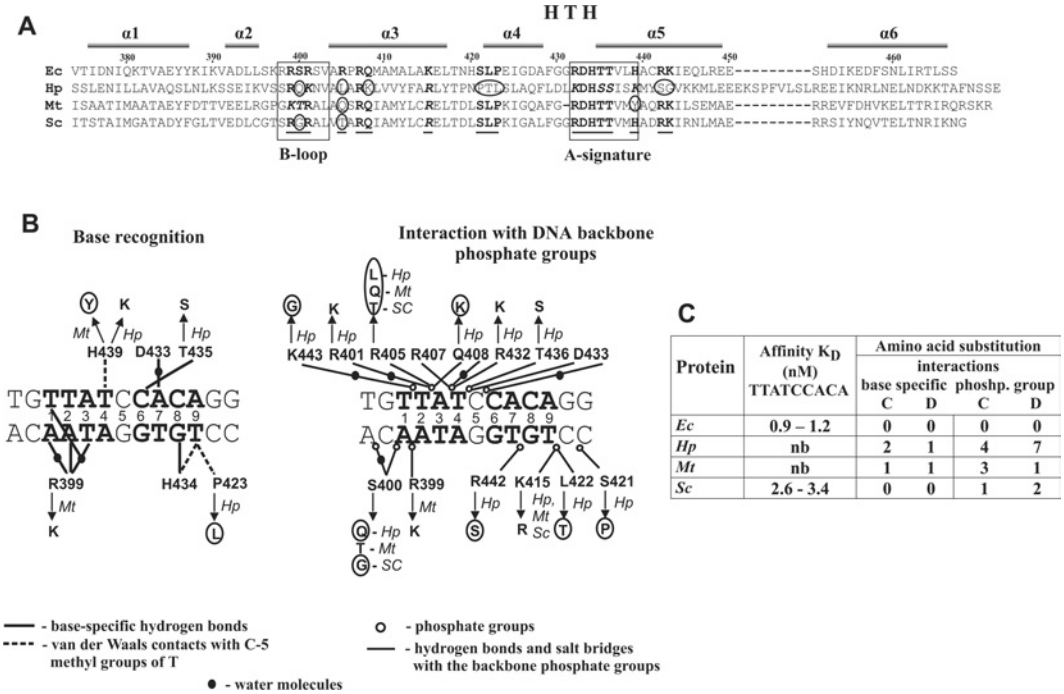
of the *S. coelicolor* DnaA protein does not contain any substitutions in amino acids that are directly involved in base recognition (Figure 7). Thus it explains why the *S. coelicolor* DnaA protein, in contrast with that of *H. pylori* and *M. tuberculosis*, binds the *E. coli* R1 DnaA box with higher affinity than its own DnaA box. In contrast with T*, C* at the third position of *MtSc*DnaA box is probably unable to form a water-mediated hydrogen bond with the highly conserved Arg^{399} (present in *S. coelicolor*, but replaced by a closely related residue, lysine, in *M. tuberculosis*) [16]. Thus the third position of the DnaA box may also contribute to the high-affinity binding. Interesting differences between the *E. coli* and the analysed proteins also concern residues that are directly involved in the interactions with DNA backbone phosphate groups (Figure 7), e.g. the triad PTL (Pro-Thr-Leu) present in the *H. pylori* binding domain probably contributes to the lower specificity of the protein; all three residues show significant divergence from the consensus, and threonine is the most divergent residue of all [24].

Analysis of the interactions between the DnaA proteins and two DnaA boxes demonstrated that, besides the sequences of DnaA boxes, their relative spacing and orientation affect binding of the proteins, particularly those of *H. pylori* and *S. coelicolor*. The *H. pylori* DnaA protein exhibits the highest affinity for the two DnaA boxes arranged 'head-to-tail' (2 bp between); such arrangement is repeated twice within the *H. pylori oriC* region (Figures 1 and 3). The *S. coelicolor* DnaA protein prefers a 'head-to-head' (3 bp between) arrangement that is present not only within its *oriC* region, but also in the promoter region of the *dnaA* gene (its expression is autoregulated) [35,36,41]. Insertion of exactly 10 bp (B-DNA has 10 bp per turn of helix) between the two *S. coelicolor* DnaA boxes practically did not change the affinity of the *S. coelicolor* DnaA protein, while such a separation between the two *H. pylori* DnaA boxes nearly completely abolished binding of the *H. pylori* DnaA protein (Figure 3). Thus in contrast with the *H. pylori* DnaA, the *S. coelicolor* DnaA protein could contact widely spaced (to some extent) DnaA boxes. Interestingly, *H. pylori* and *S. coelicolor* DnaA proteins possess the shortest (66 amino acids) and the longest (253 amino acids) domain II respectively. Domain II does not contain any

Table 2 Replication activity of the analysed *oriC* regions in *E. coli* and *S. coelicolor*

For details of transformation efficiency, see the Materials and methods section. nd, not determined. pEX4 is an *E. coli*/*Streptomyces* shuttle vector.

Plasmid	Transformation efficiency in *E. coli* (number of transformants per μg of DNA)		Minichromosome	Replication activity in *S. coelicolor** (no. of transformants per μg of DNA)
	WM1838 *polA1* (30 °C)	WM 1785 *polA*$^+$ (37 °C)		
pBR322_*HporiC*	0	9.1×10^5	p*HporiC*	nd
pBR322_*EcoriC*	14.9×10^5	3.8×10^5	p*EcoriC*	–
pBR322_*MtoriC*	0	3.2×10^5	p*MtoriC*	–
pBR322_*ScoriC*	0	2.2×10^5	p*ScoriC*	+(~100)
pBR322	0	3.9×10^5	pXE4	+(10^5–10^6)



C

Protein	Affinity K_D (nM) TTATCCACA	Amino acid substitution interactions base specific C	D	phoshp. group C	D
Ec	0.9 – 1.2	0	0	0	0
Hp	nb	2	1	4	7
Mt	nb	1	1	3	1
Sc	2.6 - 3.4	0	0	1	2

Figure 7 Comparison of the amino acid sequences of the DNA-binding domains of DnaA from *H. pylori* (*Hp*), *E. coli* (*Ec*), *M. tuberculosis* (*Mt*) and *S. coelicolor* (*Sc*)

(**A**) Protein alignment of DNA-binding domains. Grey bars above the sequence symbolize α-helices (numbered α1–α6, according to the crystal structure [16]). The positions of the HTH motif, the signature sequence motif and the basic loop are indicated [16]. Residues involved directly in DNA binding are underlined [16]. For Hp, Mt and Sc, residues identical with or similar to the *E. coli* residues involved directly in DNA binding are indicated respectively by bold or italic bold letters; non-closely related residues are encircled. (**B**) Schematic diagram summarizing the DNA (R1 DnaA box) contacts by the *E. coli* DNA-binding domain (according to Fujikawa et al. [16]). The essential base pairs in the R1 DnaA box are bold [16]. Similar and divergent (encircled) residues of other analysed organisms are indicated. (**C**) Comparison of affinities of the analysed proteins toward the *E. coli* R1 DnaA box. Dissociation constants (K_d) for the DnaA protein interaction with the *E. coli* R1 DnaA box were measured in gel retardation and SPR assays (for details see [34]). The left-hand part of the Table shows the number of amino acid substitutions in the fragment of the binding domain that is involved directly in DnaA box binding; residues involved directly in base recognition (base specific) and responsible for interaction with DNA backbone phosphate groups (phosph. group) are specified separately (see also **B**). C, conservative substitution; D, divergent substitution; nb, not bound.

relevant secondary structural motif and is considered as a flexible linker that connects the N-terminal domain involved in dimerization (domain I) with the C-terminal part of the DnaA protein responsible for ATPase activity (domain III) and DNA binding (domain IV) [9]. The presence of the long flexible domain II allows the *S. coelicolor* DnaA protein to bind separated DnaA boxes, while the efficient binding of the *H. pylori* DnaA is limited to the closely spaced DnaA boxes, presumably because of the presence of the short domain II.

The relative orientation and spacing of paired DnaA boxes do not significantly affect the binding of *E. coli* and *M. tuberculosis* DnaA proteins (domain II of *E. coli* and *M. tuberculosis* DnaA proteins consist of 77 amino acids and 83 amino acids respectively; Figure 1). In the case of *E. coli*, most DnaA boxes in the fragments analysed, i.e. in *oriC*, are strong boxes, and therefore it is not surprising that DnaA binds individually to them with no apparent co-operativity. However, detailed SPR analysis demonstrated that also for *E. coli* DnaA, the spatial arrangement of the DnaA boxes has an effect on the binding affinity of this protein [9].

The architecture of the nucleoprotein complexes: the *oriC* regions are optimally adjusted to their cognate DnaA proteins

The number and positions of the DnaA boxes determine the architecture and composition of the initiator complexes. Successive binding to sites with different affinities generates a particular order of assembly. The arrangement of DnaA boxes generates protein–DNA and protein–protein interactions that are unique to a given organism. Binding of the DnaA proteins from *E. coli* and *H. pylori* to their cognate *oriC* regions proceeds sequentially. However, the *E. coli* DnaA protein exhibits much higher affinity for its individual DnaA boxes than the *H. pylori* DnaA protein does for its own DnaA boxes (Figure 2). In contrast with the *H. pylori oriC* region, the DnaA boxes from the *E. coli oriC* are separated by considerably long spacers (see Figure 1). The lower affinity of the *H. pylori* DnaA protein towards DnaA boxes is presumably compensated by the 'compact' arrangement of the DnaA boxes (Figure 1). In the case of *E. coli* DnaA, binding starts at box number R4, and then other boxes are bound with no apparent co-operativity [21]. However, in the last step of initiation of replication, the DnaA box R1 serves as an anchor for DnaA molecules that are engaged in the unwinding reaction [9,11,12,42].

The preference of the *S. coelicolor* and *M. tuberculosis* DnaA proteins in binding concertedly several separate DnaA boxes distinguishes these complexes from those of *E. coli* and *H. pylori*. In *S. coelicolor*, there is only one DnaA box with the consensus sequence (box **6**) and none in *M. tuberculosis*. Consequently, only a few of the 19 DnaA boxes from the *oriC* region exhibit specific binding of the DnaA protein, while none of the 13 DnaA boxes from the *M. tuberculosis oriC* region is recognized individually by its own DnaA protein. At least two 'weak' DnaA boxes are required for specific efficient binding [25,30,34]. Thus it implies co-operativity for binding to origins with low-affinity DnaA

boxes. Presumably due to the high GC pressure exerted during the course of *Mycobacterium* and *Streptomyces* evolution, the *oriC* regions of *S. coelicolor* and *M. tuberculosis* have been changed; the presence of 'weak' GC-rich DnaA boxes within the *S. coelicolor* and *M. tuberculosis oriC* regions is compensated by their abundance (see Figure 1). The *S. coelicolor* nucleoprotein complex is more intricate than other analysed initiation complexes [35]. The *S. coelicolor* DnaA protein, owing to the presence of long domain II, is able to bind widely spaced DnaA boxes and to bend the short spacer that links two clusters of DnaA boxes. The SfiI probing demonstrated that the *S. coelicolor oriC* is exclusively bent only by its own DnaA protein. The other proteins analysed containing shorter domain II are not able to bend the *S. coelicolor oriC* region.

Our experiments demonstrated that the *oriC* regions, particularly those of *H. pylori* and *S. coelicolor*, are optimally adjusted to their cognate DnaA proteins. During evolution, the structure of the *oriC* regions changed, particularly the sequence and the length of the spacers that link DnaA boxes. Alterations in spacer lengths caused changes in the spatial arrangement of the DnaA boxes. The sequence of DnaA boxes also evolved differently for low- and high-GC-rich organisms such as *H. pylori* and *S. coelicolor* respectively. The weak interactions between a single DnaA box and DnaA protein are presumably compensated by the arrangement of DnaA boxes and/or by an abundance of DnaA boxes, e.g. a pair of closely spaced *H. pylori* DnaA boxes are bound by its cognate DnaA protein with the highest affinity. Thus, during the evolution of a given organism, the two key elements of initiation of replication, DnaA protein and *oriC* region, were tuned to optimal interaction. Our *in vivo* functional analysis of replication origins indicates the host specificity of *oriC* regions. The results corroborate earlier observations; the analysis of the hybrid *B. subtilis*/*E. coli* replication origin suggested that the species specificity resides in the DnaA box part of the origin, probably in the spatial arrangement of DnaA boxes [43].

The results of the present study indicate that the primary functions of multiple DnaA boxes are to determine the positioning and order of assembly of the DnaA molecules. Gradual transition from the sequence-specific binding of the DnaA protein to binding through co-operative protein–protein interactions seems to be a common conserved strategy to generate oligomeric initiator complexes that are bound to multiple sites within the chromosomal, plasmid [44] and viral origins [45–48].

This work was supported by the Ministry of Scientific Research and Information Technologies Society (grant 3 P04A 079 22). D.J. gratefully acknowledges financial support received from the EC (HPMF-CT-2002-01676).

REFERENCES

1 Kornberg, A. and Baker, T. A. (1992) DNA Replication, W. H. Freeman & Co., New York
2 Baker, T. A. and Bell, S. P. (1998) Polymerases and the replisome: machines within machines. Cell **92**, 295–305
3 Matsunaga, F., Forterre, P., Ishino, Y. and Myllykallio, H. (2001) *In vivo* interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. Proc. Natl. Acad. Sci. U.S.A. **98**, 11152–11157
4 Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., Bernander, R. and Bell, S. D. (2004) Identification of two origins of replication in the single chromosome of the archaeon *Sulfolobus solfataricus*. Cell **116**, 25–38
5 Bell, S. P. (2002) The origin recognition complex: from simple origins to complex functions. Genes Dev. **16**, 659–672
6 Bell, S. P. and Dutta, A. (2002) DNA replication in eukaryotic cells. Annu. Rev. Biochem. **71**, 333–374
7 Kelman, L. M. and Kelman, Z. (2003) Archaea: an archetype for replication initiation studies? Mol. Microbiol. **48**, 605–615
8 Giraldo, R. (2003) Common domains in the initiators of DNA replication in Bacteria, Archaea and Eukarya: combined structural, functional and phylogenetic perspectives. FEMS Microbiol. Rev. **26**, 533–554
9 Messer, W. (2002) The bacterial replication initiator DnaA: DnaA and *oriC*, the bacterial mode to initiate DNA replication. FEMS Microbiol. Rev. **26**, 355–374
10 Erzberger, J. P., Pirruccello, M. M. and Berger, J. M. (2002) The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. EMBO J. **21**, 4763–4773
11 Messer, W., Blaesing, F., Jakimowicz, D., Krause, M., Majka, J., Nardmann, J., Schaper, S., Seitz, H., Speck, C., Weigel, C. et al. (2001) Bacterial replication initiator DnaA: rules for DnaA binding and roles of DnaA in origin unwinding and helicase loading. Biochimie **83**, 5–12
12 Speck, C. and Messer, W. (2001) Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. EMBO J. **20**, 1469–1476
13 Messer W., Blaesing, F., Majka, J., Nardmann, J., Schaper, S., Schmidt, A., Seitz, H., Speck, C., Tungler, D., Wegrzyn G. et al. (1999) Functional domains of DnaA proteins. Biochimie **81**, 819–825
14 Weigel, C., Schmidt, A., Seitz, H., Tungler, D., Welzeck, M. and Messer, W. (1999) The N-terminus promotes oligomerization of the *Escherichia coli* initiator protein DnaA. Mol. Microbiol. **34**, 53–66
15 Simmons, L. A., Felczak, M. and Kaguni, J. M. (2004) DnaA protein of *Escherichia coli*: oligomerization at the *E. coli*. chromosomal origin is required for initiation and involves specific N-terminal amino acids. Mol. Microbiol. **49**, 849–858
16 Fujikawa, N., Kurumizaka, H., Nureki, O., Terada, T., Shirouzu, M., Katayama, T. and Yokoyama, S. (2003) Structural basis of replication origin recognition by the DnaA protein. Nucleic Acids Res. **31**, 2077–2086
17 Mackiewicz, P., Zakrzewska-Czerwińska, J., Zawilak, A., Dudek, M. R. and Cebrat, S. (2004) Where does bacterial replication start? Rules for predicting the *oriC* region. Nucleic Acids Res. **32**, 3781–3791
18 Woelker, B. and Messer, W. (1993) The structure of the initiation complex at the replication origin, *oriC*, of *Escherichia coli*. Nucleic Acids Res. **21**, 5025–5033
19 Hopwood, D. A., Bibb, M. J., Chater, K. F., Kieser, T., Bruton, C. J., Kieser, H. M., Lydiate, D. J., Smith, C. P., Ward, J. M. and Schrempf, H. (1985) Genetic Manipulation of *Streptomyces*: a Laboratory Manual, The John Innes Foundation, Norwich
20 Zawilak, A. (2003) Characterization of the key elements of initiation of *Helicobacter pylori* chromosome replication, Ph.D. Thesis, Ludwik Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Wrocław
21 Weigel, C., Schmidt, A., Ruckert, B., Lurz, R. and Messer, W. (1997) DnaA protein binding to individual DnaA boxes in the *Escherichia coli* replication origin, *oriC*. EMBO J. **16**, 6574–6583
22 Zakrzewska-Czerwińska, J., Majka, J. and Schrempf, H. (1995) Minimal requirements of the *Streptomyces lividans* 66 *oriC* region and its transcriptional and translational activities. J. Bacteriol. **177**, 4765–4771
23 Krause, M., Rückert, B., Lurz, R. and Messer, W. (1997) Complexes at the replication origin of *Bacillus subtilis* with homologous and heterologous DnaA protein. J. Mol. Biol. **274**, 365–380
24 Zawilak, A., Durrant, M. C., Jakimowicz, P., Backert, S. and Zakrzewska-Czerwińska, J. (2003) DNA binding specificity of the replication initiator protein, DnaA from *Helicobacter pylori*. J. Mol. Biol. **334**, 933–947
25 Zawilak, A., Kois, A., Konopa, G., Smulczyk-Krawczyszyn, A. and Zakrzewska-Czerwińska, J. (2004) *Mycobacterium tuberculosis* DnaA initiator protein: purification and DNA binding requirements. Biochem. J. **382**, 247–252
26 Majka, J., Messer, W., Schrempf, H. and Zakrzewska-Czerwińska, J. (1997) Purification and characterization of the *Streptomyces lividans* initiator protein DnaA. J. Bacteriol. **179**, 2426–2432
27 Parada, C. A. and Marinas, K. J. (1991) Mechanism of DNA A protein-dependent pBR322 DNA replication: DNA A protein-mediated *trans*-strand loading of the DNA B protein at the origin of pBR322 DNA. J. Biol. Chem. **266**, 18895–18906
28 Carey, J. (1991) Gel retardation. Methods Enzymol. **208**, 103–117
29 Majka, J., Jakimowicz, D., Messer, W., Schrempf, H., Lisowski, M. and Zakrzewska-Czerwińska, J. (1999) Interactions of the *Streptomyces lividans* initiator protein DnaA with its target. Eur. J. Biochem. **260**, 325–335
30 Jakimowicz, D., Majka, J., Messer, W., Speck, C., Fernandez, M., Martin, M. C., Sanchez, J., Schauwecker, F., Keller, U., Schrempf, H. and Zakrzewska-Czerwińska, J. (1998) Structural elements of the *Streptomyces oriC* region and their interactions with the DnaA protein. Microbiology **144**, 1281–1290
31 Qin, M. H., Madiraju, M. V. and Rajagopalan, M. (1999) Characterization of the functional replication origin of *Mycobacterium tuberculosis*. Gene **233**, 121–130
32 Rajagopalan, M., Qin, M.-H., Nash, D. R. and Madiraju, M. V. V. S. (1995) *Mycobacterium smegmatis dnaA* region and autonomous replication activity. J. Bacteriol. **177**, 6527–6535

33 Zawilak, A., Cebrat, S., Mackiewicz, P., Krol-Hulewicz, A., Jakimowicz, D., Messer, W., Gosciniak, G. and Zakrzewska-Czerwińska, J. (2001) Identification of a putative chromosomal replication origin from *Helicobacter pylori* and its interaction with the initiator protein DnaA. Nucleic Acids Res. **29**, 2251–2259
34 Majka, J., Zakrzewska-Czerwińska, J. and Messer, W. (2001) Sequence recognition, cooperative interaction, and dimerization of the initiator protein DnaA of *Streptomyces*. J. Biol. Chem. **276**, 6243–6252
35 Jakimowicz, D., Majka, J., Konopa, G., Wegrzyn, G., Messer, W., Schrempf, H. and Zakrzewska-Czerwińska, J. (2000) Architecture of the *Streptomyces lividans* DnaA protein–replication origin complexes. J. Mol. Biol. **298**, 351–364
36 Zakrzewska-Czerwińska, J., Nardmann, J. and Schrempf, H. (1994) Inducible transcription of the *dnaA* gene from *Streptomyces lividans* 66. Mol. Gen. Genet. **242**, 440–447
37 Margulies, C. and Kaguni, J. M. (1996) Ordered and sequential binding of DnaA protein to *oriC*, the chromosomal origin of *Escherichia coli*. J. Biol. Chem. **271**, 17035–17040
38 Schaper, S. and Messer, W. (1995) Interaction of the initiator protein DnaA of *Escherichia coli* with its DNA target. J. Biol. Chem. **270**, 17622–17626
39 Wentzell, L. M. and Halford, S. E. (1998) DNA looping by the *Sfi*I restriction endonuclease. J. Mol. Biol. **281**, 433–444
40 Langer, U., Richter, S., Roth, A., Weigel, C. and Messer, W. (1996) A comprehensive set of DnaA-box mutations in the replication origin, *oriC*, of *Escherichia coli*. Mol. Microbiol. **21**, 301–311
41 Jakimowicz, D., Majka, J., Lis, B., Konopa, G., Wegrzyn, G., Messer, W., Schrempf, H. and Zakrzewska-Czerwińska, J. (2000) Structure and regulation of the *dnaA* promoter region in three *Streptomyces* species. Mol. Gen. Genet. **262**, 1093–1102
42 McGarry, K. C., Ryan, V. T., Grimwade, J. E. and Leonard, A. C. (2004) Two discriminatory binding sites in the *Escherichia coli* replication origin are required for DNA strand opening by initiator DnaA–ATP. Proc. Natl. Acad. Sci. U.S.A. **101**, 2811–2816
43 Seitz, H., Welzeck, M. and Messer, W. (2001) A hybrid bacterial replication origin. EMBO Rep. **2**, 1003–1006
44 Brendler, T. G., Abeles, A. L., Reaves, L. D. and Austin, S. J. (1997) The iteron bases and spacers of the P1 replication origin contain information that specifies the formation of a complex structure involved in initiation. Mol. Microbiol. **23**, 559–567
45 Tjian, R. (1978) The binding site on SV40 DNA for a T antigen-related protein. Cell **13**, 165–179
46 Chen, G. and Stenlund, A. (1998) Characterization of the DNA-binding domain of the bovine papillomavirus replication initiator E1. J. Virol. **72**, 2567–2576
47 Chen, G. and Stenlund, A. (2001) The E1 initiator recognizes multiple overlapping sites in the papillomavirus origin of DNA replication. J. Virol. **75**, 292–302
48 Chen, G. and Stenlund, A. (2002) Sequential and ordered assembly of E1 initiator complexes on the papillomavirus origin of DNA replication generates progressive structural changes related to melting. Mol. Cell. Biol. **22**, 7712–7720

Received 20 January 2005/14 March 2005; accepted 24 March 2005
Published as BJ Immediate Publication 24 March 2005, DOI 10.1042/BJ20050143